Promotion of Open Science in Requirements Engineering: Leveraging the Open Research Knowledge Graph for FAIR Scientific Information

Oliver Karras*, Alessio Ferrari†, Davide Fucci‡, and Davide Dell'Anna§
*TIB - Leibniz Information Centre for Science and Technology, Germany, Email: oliver.karras@tib.eu
†CNR-ISTI, Italy, Email: alessio.ferrari@isti.cnr.it
‡Blekinge Institute of Technology, Sweden, Email: davide.fucci@bth.se
§Utrecht University, The Netherlands, Email: d.dellanna@uu.nl

***Abstract*—[Background.] Despite improved digital access to publications as digitized artifacts, they remain document-based and often behind paywalls, impeding open science. Researchers must push beyond the established boundary of publications as digitized documents. Open science infrastructures support them in organizing and (re-)using publications and their information so that they are Findable, Accessible, Interoperable, and Reusable (FAIR) for humans and machines in the long term. The Open Research Knowledge Graph (ORKG) is one sustainably governed infrastructure for FAIR scientific information, with successful use cases in requirements engineering (RE). [Objective.] This tutorial aims to familiarize the participants with open science and empower them to leverage the ORKG for FAIR scientific information. [Method.] The half-day tutorial consists of three sessions: 1) A short theoretical introduction to open science and the ORKG regarding their importance, benefits, and incentives, 2) A practical session with hands-on exercises for learning skills and practical experiences to leverage the ORKG, and 3) A feedback session for reflection. [Results.] The tutorial raises awareness for open science in RE, introduces the ORKG, fosters networking, and, in the best case, establishes future collaborations. The participants become familiar with open science and the ORKG by learning skills and gaining practical experience to leverage the ORKG themselves. [Conclusions.] The transition away from digitized documents and towards FAIR scientific information is a long-term endeavor. We must gradually sensitize researchers to this transition while guiding and empowering them to leverage existing solutions as an integral part of their work. The tutorial pursues this endeavor to promote open science in RE.**

***Index Terms*—Open science, tutorial, knowledge graph, infrastructure, FAIR, scientific information, requirements engineering**

I. Motivation and Objectives

With the growing number of publications (approx. seven million per year) [1], open science has become increasingly important [2]. Research communities face the challenge of keeping pace with the rapidly growing and constantly evolving state of the art to make targeted contributions. Given the importance of open science, researchers need to be familiarized and guided with this topic to push beyond the established boundary of publications as digitized documents. The next step in the digital transformation of scientific information is its long-term Findable, Accessible, Interoperable, and Reusable (FAIR) representation [3]. This change requires researchers to transition to using open science infrastructures that support them in organizing and (re-)using publications and their FAIR scientific information to make them understandable, processable, and usable by humans and machines[4].

The *Open Research Knowledge Graph* (ORKG) is a ready-to-use and sustainably governed open science infrastructure for the production, curation, (re-)use, and publication of FAIR scientific information [5]–[7]. The ORKG comprises the technical infrastructure, services, interfaces, and tools for researchers to collect, extract, analyze, and publish FAIR scientific information collaboratively on any research domain, problem, or question. Besides various successful use cases of the ORKG in different research domains [7], there are use cases in requirements engineering (RE) [8]–[11]. One of these use cases received the Best Paper Award of the 17th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement 2023 [9] and the badge "Reusable" for the associated artifact by the Artifact Evaluation Track of the 32nd IEEE International Requirements Engineering Conference 2024 [11].

Several conferences, i.a., *SEMANTiCS'22*, *SEMANTiCS'23* & *SEMANTiCS'24*, *ISWC'23* & *ISWC'24*, and the journal *ing.grid* already include the ORKG as a part of their open science initiatives. In 2023, *REFSQ'24* and the *7th NLP4RE workshop* have joined this movement. REFSQ'24 included the ORKG in its open science track, which *REFSQ'25* continues. The 7th NLP4RE workshop considered using the ORKG for the so-called ID-Card that provides a structured summary of research publications, emphasizing replication-relevant information [12]. The ID-Card was presented last year in the tutorial *Towards Facilitating Replication in NLP4RE* at RE'23, where a talk on the ORKG was spontaneously requested due to its potential to increase replicability in science [13]. These developments motivated us to launch this *tutorial* at RE'24. Below, we summarize our motivation and objectives.

Summary of Motivation:
1) Growing interest and need for awareness of open science, its importance, benefits, and incentives in RE.
2) Need for guidance and empowerment of researchers to transition to using open science infrastructures, such as the ORKG, as an integral part of their work.
3) Supporting the open science initiatives of REFSQ'24, REFSQ'25, and the 7th NLP4RE workshop to promote open science with the ORKG in the wider RE community.

Objectives:
1) Raising awareness and educating researchers about open science, its importance, benefits, and incentives for themselves and the RE community.
2) Introducing the ORKG to the wider RE community and empowering researchers through hands-on exercises to learn skills and practical experiences to leverage the ORKG themselves for their own use cases.
3) Joining the movement of REFSQ'24, REFSQ'25, and the 7th NLP4RE workshop by familiarizing researchers with the ORKG at an early stage, thus establishing the ORKG as an integral part of their work in the near future.

II. Target Audience

The tutorial is aimed at anyone interested in open science and who wants to learn skills and gain practical experience to contribute to open science in RE by leveraging the ORKG for FAIR scientific information.

III. Outline of the Tutorial

In Table I, we present the outline of the tutorial. We split the tutorial into a theoretical session (25 min), a practical session (110 min), and a feedback session (30 min). All slides and materials of the tutorial are publicly available on Zenodo [14].

The theoretical session serves to familiarize all participants with open science and the ORKG so that they can actively participate in the tutorial and perform the exercises themselves.

The practical session consists of two exercises: 1) Using SciKGTeX, a tool from the ORKG, to create a FAIR-annotated publication and 2) Using the ORKG to create an ORKG comparison. Both exercises align with the open science initiative of REFSQ'24 and REFSQ'25. While the open science track of REFSQ'24 encouraged authors to use these workflows, the open science track of REFSQ'25 even launched an open science competition with two challenges, one of which relates to one of the exercises. For both exercises, the tutorial interactively walks through all the necessary steps so that researchers learn the skills and practical experiences required to leverage SciKGTeX and the ORKG themselves to promote open science in RE.

In the first exercise, the participants use the LaTeX package SciKGTeX [15], which won the Vannevar Bush Best Paper Award at ACM/IEEE Joint Conference on Digital Libraries 2023. This tool enables authors to annotate FAIR scientific information at the time of writing a publication and embed it in the generated PDF so that it is long-term available and can even be uploaded to the ORKG. This exercise is related to the first challenge of the open science competition of the open science track of REFSQ'25. The accepted publication, that is best annotated with SciKGTeX, will receive the Best ORKG Annotation Award. We want to support this challenge by empowering researchers to annotate their publications with SciKGTeX and upload them to the ORKG to contribute to open science in RE and, in the best case, win the award.

In the second exercise, the participants use the ORKG to create an ORKG comparison. Overall, they learn to add publications to the ORKG and describe them regarding their scientific information. Based on these publications, the participants create and publish an ORKG comparison. This exercise is related to the second challenge of the open science competition of the open science track of REFSQ'25. The accepted paper, that is enriched with the best ORKG comparison, will receive the Best ORKG Comparison Award. We also want to support this challenge by empowering researchers to leverage the ORKG themselves so that they can develop their own use cases and, in the best case, win the award.

In the feedback session, we reflect on the tutorial with the participants to improve it for future iterations so that it better aligns with the needs of researchers in RE and beyond.

IV. Tutorial History

While this tutorial is proposed for the first time, the organizers have extensive experience on its topics open science,

Table I: Outline of the Tutorial

Session	Time	Table of Content	Style	Speaker
Theoretical	09:00 - 09:25	1. Welcome (5 min)	Presentation	All organizers
		2. Introduction to open science in RE (10 min)	Presentation	Alessio Ferrari
		3. Introduction to the ORKG (10 min)	Presentation	Oliver Karras
Practical	09:25 - 10:15	4. Create a FAIR-annotated publication for the ORKG (50 min)	Exercise	Oliver Karras
		4.1 Set up an Overleaf project for an exemplary publication	Sub-exercise	All organizers
		4.2 Use the LaTex package SciKGTeX to annotate the publication	Sub-exercise	All organizers
		4.3 Generate PDF with embedded FAIR scientific information	Sub-exercise	All organizers
		4.4 Optional: Upload the FAIR-annotated publication to the ORKG	Sub-exercise	All organizers
Break	10:15 - 10:45	Coffee break		
Practical	10:45 - 11:45	5. Use the ORKG based on a RE use case (60 min)	Exercise	Oliver Karras
		5.1 Add an exemplary publication to the ORKG	Sub-exercise	All organizers
		5.2 Describe the scientific information of the publication in the ORKG	Sub-exercise	All organizers
		5.3 Create an ORKG comparison of the publications added by participants	Sub-exercise	All organizers
		5.4 Publish the created ORKG comparison as a citable digital artifact	Sub-exercise	All organizers
		5.5 Optional: Create visualizations for the created ORKG comparison	Sub-exercise	All organizers
		5.6 Optional: Retrieve the information with the SPARQL endpoint	Sub-exercise	All organizers
Feedback	11:45 - 12:15	6. Reflection of the tutorial with the participants (25 min)	Discussion	All organizers
		7. Farewell and closing (5 min)	Presentation	All organizers

RE, and ORKG. The first organizer regularly offers talks and interactive workshops on the ORKG. For example, he gave talks about the ORKG and its use cases from various research domains, including RE, at *5th International Workshop on Crowd-based Requirements Engineering 2021* [8], *Scientific Data Conference 2021*, *International Council for Scientific and Technical Information (ICSTI) Exchanges: Knowledge Representation and Data Visualization 2022*, *NFDI4Ing - Community Meeting CC-43 2022*, *67th Helmholtz Open Science Online Seminar 2023* [5], and *1st NFDI4Energy Conference 2024* [16]. In addition, he organized an interactive session on the ORKG at the *5th International Workshop on Crowd-based Requirements Engineering 2021* and conducted two interactive workshops on the ORKG at the *Open Science Festival 2022* and the *Open Science Festival 2023*. The other organizers have conducted another similar tutorial *Towards Facilitating Replication in NLP4RE* at RE'23 and some of them were the co-chairs of the open science track at *REFSQ'24*.

V. Biographies of the Organizers

Oliver Karras is a researcher, data scientist, and lecturer working at TIB – Leibniz Information Centre for Science and Technology. His research includes developing use cases of the ORKG and supporting researchers in developing their use cases for FAIR scientific information. He frequently serves as a member of program and organizing committees of conferences, i.a., CIKM, IRI, REFSQ, and RE. He was local chair of RE'23 and is one of the open science co-chairs of REFSQ'25.

Alessio Ferrari is a research scientist at Consiglio Nazionale delle Ricerche - Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (CNR-ISTI), Italy. His research focuses on NLP applied to RE, and on empirical studies on requirements elicitation interviews. He has been part of the RE and ICSE program committee, local organization co-chair of REFSQ'20, program committee co-chair of REFSQ'23, and will be program committee co-chair of RE'25. He is one of the creators and co-organizers of the NLP4RE workshop.

Davide Fucci is an associate professor at the Department of Software Engineering, at the Blekinge Institute of Technology in Sweden. His research interests in RE combines NLP, affective computing, and empirical research. He was the organizer of the 3rd NLP4RE workshop co-located with REFSQ'20, as well as the 1st and 2nd AffectRE workshop co-located with RE'18 and RE'19.

Davide Dell'Anna is an assistant professor at the Department of Information and Computing Sciences of Utrecht University, The Netherlands. His research focuses on methodologies and solutions for AI systems that cooperate with humans. He regularly serves as a program committee member for international conferences such as RE, REFSQ, AAMAS, IJCAI, ECAI. He is one of the co-organizers of REFSQ'18, NLP4RE'22 and NLP4RE'24.

Data Availability

We provide all slides and materials from the tutorial publicly available on Zenodo [14] so that everyone can repeat the theoretical presentations independently and carry out the practical exercises themselves at any time.

Acknowledgment

The authors thank the Federal Government, the Heads of Government of the Länder, as well as the Joint Science Conference (GWK), for their funding and support within the NFDI4Ing, NFDI4DataScience, and NFDI4Energy consortia. This work was funded by the German Research Foundation (DFG) project numbers 442146713, 460234259, and 501865131 and by the European Research Council for the project ScienceGRAPH (Grant agreement ID: 819536). This work was partially supported by the KKS foundation through the S.E.R.T. Research Profile project at Blekinge Institute of Technology.

References

[1] M. Fire and C. Guestrin, "Over-Optimization of Academic Publishing Metrics: Observing Goodhart's Law in Action," *GigaScience*, vol. 8, no. 6, 2019.

[2] D. Mendez, D. Graziotin, S. Wagner, and H. Seibold, "Open Science in Software Engineering," in *Contemporary Empirical Methods in Software Engineering*. Springer, 2020.

[3] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton *et al.*, "The FAIR Guiding Principles for Scientific Data Management and Stewardship," *Scientific Data*, vol. 3, no. 1, 2016.

[4] S. Auer, D. A. C. Barone, C. Bartz, E. G. Cortes, M. Y. Jaradeh, O. Karras *et al.*, "The SciQA Scientific Question Answering Benchmark for Scholarly Knowledge," *Nature Scientific Reports*, vol. 13, no. 7240.

[5] S. Auer, M. Stocker, O. Karras, A. Oelen, J. D'Souza, and A.-L. Lorenz, "Organizing Scholarly Knowledge in the Open Research Knowledge Graph: An Open-Science Platform for FAIR Scholarly Knowledge," in *Proceedings of the Conference on Research Data Infrastructure*, 2023.

[6] M. Stocker, A. Oelen, M. Y. Jaradeh, M. Haris, O. A. Oghli *et al.*, "FAIR Scientific Information with the Open Research Knowledge Graph," *FAIR Connect*, vol. 1, no. 1, 2023.

[7] S. Auer, V. Ilangovan, M. Stocker, S. Tiwari, and L. Vogt, *Open Research Knowledge Graph*. Cuvillier Verlag, 2024.

[8] O. Karras, E. C. Groen, J. A. Khan, and S. Auer, "Researcher or Crowd Member? Why not both! The Open Research Knowledge Graph for Applying and Communicating CrowdRE Research," in *29th International Requirements Engineering Conference Workshops*. IEEE, 2021.

[9] O. Karras, F. Wernlein, J. Klünder, and S. Auer, "Divide and Conquer the EmpiRE: A Community-Maintainable Knowledge Graph of Empirical Research in Requirements Engineering," in *International Symposium on Empirical Software Engineering and Measurement (ESEM)*. IEEE, 2023.

[10] ——, "KG-EmpiRE: A Community-Maintainable Knowledge Graph of Empirical Research in Requirements Engineering," in *Software Engineering 2024*, 2024.

[11] O. Karras, "KG-EmpiRE: A Community-Maintainable Knowledge Graph for a Sustainable Literature Review on the State and Evolution of Empirical Research in Requirements Engineering," in *32nd International Requirements Engineering Conference (RE)*. IEEE, 2024.

[12] S. Abualhaija, F. B. Aydemir, F. Dalpiaz, D. Dell'Anna, A. Ferrari *et al.*, "Replication in Requirements Engineering: The NLP for RE Case," *ACM Transactions on Software Engineering and Methodology*, 2024.

[13] H. Hussein, K. E. Farfar, A. Oelen, O. Karras, and S. Auer, "Increasing reproducibility in science by interlinking semantic artifact descriptions in a knowledge graph," in *Leveraging Generative Intelligence in Digital Libraries: Towards Human-Machine Collaboration*. Springer, 2023.

[14] O. Karras, A. Ferrari, D. Fucci, and D. Dell'Anna, "Supplementary Materials of the Tutorial: "Promotion of Open Science in Requirements Engineering - Leveraging the Open Research Knowledge Graph for FAIR Scientific Information"." Zenodo, 2024, 10.5281/zenodo.11354706.

[15] C. Bless, I. Baimuratov, and O. Karras, "SciKGTeX - A LaTeX Package to Semantically Annotate Contributions in Scientific Publications," in *23nd ACM/IEEE Joint Conference on Digital Libraries*, 2023.

[16] O. Karras, J. Göpfert, P. Kuckertz, T. Pelser, and S. Auer, "Organizing Scientific Knowledge From Energy System Research Using the Open Research Knowledge Graph," in *1st NFDI4Energy Conference*. Zenodo, 2024.